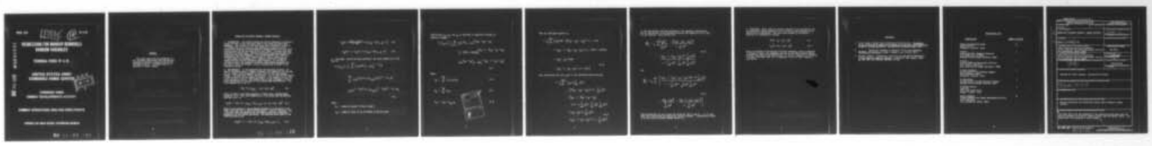
AD-A061 021 ARMY COMBINED ARMS COMBAT DEVELOPMENTS ACTIVITY FORT--ETC F/G 16/2
REGRESSION FOR MARKOV BERNOULLI RANDOM VARIABLES.(U)
APR 78 E H INSELMANN

UNCLASSIFIED CACDA-TP-4-78 NL

1 OF 1
AD
A061021



END
DATE
FILMED
1-79
DDC

APRIL 1978 LEVEL II 12 TP 4-78

REGRESSION FOR MARKOV BERNOULLI RANDOM VARIABLES

TECHNICAL PAPER TP 4-78

UNITED STATES ARMY COMBINED ARMS CENTER

COMBINED ARMS COMBAT DEVELOPMENTS ACTIVITY

COMBAT OPERATIONS ANALYSIS DIRECTORATE

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

ABSTRACT

This paper deals with the estimation of the regression coefficients when the data are sequences of Bernoulli random variables that form Markov chains. The method used is an extension of Klotz's papers.

REGRESSION FOR MARKOV BERNOULLI RANDOM VARIABLES

1. INTRODUCTION. The problem studied was that of regression on Bernoulli random variables in the case where some of the random variables were dependent. The interest in this case arose from a problem of trying to fit probability of hit curves to data generated by repeated missile simulations performed at US Army Materiel Systems Analysis Activity using tracking data from the Antitank Missile Test (ATMT). Hit/miss data were generated one second apart. Because overlapping tracking data were used, successive shots were dependent. This caused problems that seemed insurmountable until the author became aware of Klotz's papers (1) (2). In these papers the parameters of a sequence of Bernoulli dependent random variables satisfy the Markov chain property. In the case of successive shots, the assumption of Markov chain seemed reasonable and was used to solve the problem. Klotz's technique was extended to the regression problem.

2. PRELIMINARIES. In the generated data the following occurred: for several different ranges, a number of gunners (the number was not the same for all ranges) fired a sequence of shots (not all the same sequence length). The shots were fired a second apart. Let X(I,J,R) be the results of the Ith shot of the Jth gunner at range R. A hit caused X to be 1 and a miss caused it to be 0. The notation that is now introduced is that of Klotz but modified to the needs of the problem under consideration. The first probability of hit is:

$$P(R) = Pr\left\{X_{ijR}\right\} = b_0 + b_1 R + b_2 R^2 \qquad \text{Eq 1}$$

which, as shown in the above equation, is taken to be a second degree polynomial in R. Next, the probability of a hit given that the previous shot was a hit, is:

$$P_{11}(R) = \lambda(R) = Pr\left\{X_{ijR} = 1 | X_{i-1\ jR} = 1\right\} = a_0 + a_1 R + a_2 R^2 \qquad \text{Eq 2}$$

which is also taken as a second degree polynomial in R and which is the lower right hand term in the transition matrix. Clearly, equations 1 and 2 hold only when the sequences are stationary, which was a reasonable assumption for the problem considered. The remaining three terms of the transition matrix are:

$$P_{01}(R) = 1 - \lambda(R) = Pr\left\{X_{ijR} = 0 | X_{i-1\ jR} = 1\right\} \qquad \text{Eq 3}$$

$$P_{10}(R) = \frac{P(R)\left[1 - \lambda(R)\right]}{1 - P(R)} = Pr\left\{ X_{ijR} = 1 \middle| X_{i-1\ jR} = 0 \right\} \qquad \text{Eq 4}$$

$$P_{00}(R) = 1 - P_{10}(R) = Pr\left\{ X_{ijR} = 0 \middle| X_{i-1\ jR} = 0 \right\} \qquad \text{Eq 5}$$

3. **LIKELIHOOD.** Having the above machinery, the joint probability of the data is:

$$Pr\left\{ X_{ijR}\text{'s} \right\} = \underset{R}{\pi} \ \underset{j=1}{\overset{N_R}{\pi}} \left\{ P(R)^{X_{1jR}} \left[1 - P(R)\right]^{1 - X_{1jR}} \underset{i=2}{\overset{n_{jR}}{\pi}} P_{11}(R)^{X_{ijR} X_{i-1\ jR}} P_{10}(R)^{X_{ijR}(1 - X_{i-1\ jR})} P_{01}(R)^{(1 - X_{ijR}) X_{i-1\ jR}} P_{00}(R)^{(1 - X_{ijR})(1 - X_{i-1\ jR})} \right\} \qquad \text{Eq 6}$$

where:

N_R = number of gunners firing at range R

n_{jR} = number of shots by the jth gunner at the Rth range.

Substituting r_{jR}, S_{jR}, and t_{jR} as described in equations 8 through 10, equation 6 becomes:

$$Pr\left\{X_{ijR}'s\right\} = \pi_{R} \; \pi_{j=1}^{N_R} \left\{ \lambda(R)^{r_{jR}} (1 - \lambda(R))^{2(S_{jR} - r_{jR}) - t_{jR}} \left[1 - 2P(R) + \lambda(R)P(R)\right]^{(n_{jR} - 1 - 2S_{jR} + r_{jR} + t_{jR})} P(R)^{(S_{jR} - r_{jR})} \left[1 - P(R)\right]^{-(n_{jR} - 2 - S_{jR} + t_{jR})} \right\}$$

Eq 7

Where:

$$r_{jR} = \sum_{i=2}^{n_{jR}} X_{i-1\;jR} X_{ijR}$$ Eq 8

$$S_{jR} = \sum_{j=1}^{n_{jR}} X_{ijR}$$ Eq 9

$$t_{jR} = X_{1jR} + X_{n_{jR}\;jR}$$ Eq 10

Then the likelihood function is:

$$L = \sum_{R} \sum_{j=1}^{N(R)} \Big\{ r_{jR} \ln\lambda(R) + \left[2(S_{jR} - r_{jR}) - t_{jR}\right] \ln(1 - \lambda(R))$$

$$+ (n_{jR} - 1 - 2S_{jR} + r_{jR} + t_{jR}) \ln(1 - 2P(R) + \lambda(R)P(R))$$

$$+ (S_{jR} - r_{jR}) \ln P(R)$$

$$- (n_{jR} - 2 - S_{jR} + t_{jR}) \ln(1 - P(R)) \Big\} \qquad \text{Eq 11}$$

Now, substituting $\lambda(R)$ and $P_{11}(R)$ in the likelihood function one has:

$$L = \sum_{R} \sum_{j=1}^{N(R)} \Big\{ r_{jR} \ln\left(\sum_{k=0}^{2} a_k R^k\right)$$

$$+ \left[2(S_{jR} - r_{jR}) - t_{jR}\right] \ln\left(1 - \sum_{k=0}^{2} a_k R^k\right)$$

$$+ \left[n_{jR} - 1 - 2S_{jR} + r_{jR} + t_{jR}\right]$$

$$\ln\left(1 - 2\sum_{q=0}^{2} b_q R^q + \sum_{q=0}^{2} b_q R^q \sum_{k=0}^{2} a_k R^k\right)$$

$$+ (S_{jR} - r_{jR}) \ln\left(\sum_{q=0}^{2} b_q R^q\right)$$

$$- (n_{jR} - 2 - S_{jR} + t_{jR}) \ln\left(1 - \sum_{q=0}^{2} b_q R^q\right) \Big\} \qquad \text{Eq 12}$$

To find the maximum likelihood estimates of the regression coefficients, partial derivatives of the likelihood function with respect to the a's and b's are required. These partials are:

$$\frac{\partial L}{\partial a_m} = \sum_R \sum_{j=1}^{N(R)} \left(\frac{r_{jR} R^m}{\sum_{k=0}^{2} a_k R^k} - \frac{\left[2(S_{jR} - r_{jR}) - t_{jR}\right] R^m}{1 - \sum_{k=0}^{2} a_k R^k} \right.$$

Eq 13

$$\left. + \frac{(n_{jR} - 1 - 2S_{jR} + t_{jR})\, R^m \sum_{q=0}^{2} b_q R^q}{1 - 2 \sum_{q=0}^{2} b_q R^q + \sum_{q=0}^{2} b_q R^q \sum_{k=0}^{2} a_k R^k} \right)$$

and

$$\frac{\partial L}{\partial b_n} = \sum_R \sum_{j=1}^{N(R)} \left(\frac{(n_{jR} - 1 - 2S_{jR} + r_{jR} + t_{jR})(-2R^n + R^n \sum_{k=0}^{2} a_k R^k)}{1 - 2 \sum_{q=0}^{2} b_q R^q + \sum_{q=0}^{2} b_q R^q \sum_{k=0}^{2} a_k R^k} \right.$$

Eq 14

$$\left. + \frac{(S_{jR} - r_{jR}) R^n}{\sum_{q=0}^{2} b_q R^q} + \frac{(n_{jR} - 2 - S_{jR} + t_{jR})\, R^n}{1 - \sum_{q=0}^{2} b_q R^q} \right)$$

These expressions are set to zero and solved for the a's and b's. It is clear that the solutions must be obtained by iterative methods. A program was written to do this using the Newton Raphson method (3).

4. CONCLUSION. Recall that the problem discussed in the introduction was the problem of fitting probability of hit curves to data generated by repeated missile simulations. The curves were assumed to be quadratic functions of R expressed as follows:

$$P(R) = b_0 + b_1R + b_2R^2 \qquad \text{Eq 1}$$

$$P_{11}(R) = a_0 + a_1R + a_2R^2 \qquad \text{Eq 2}$$

Hence, utilization of the maximum likelihood technique, given by equations 11 through 14 above, and subsequent solution by the Newton Raphson method, provides the values of the coefficients, a's and b's, necessary to achieving a maximum likelihood "best fit" of equations 1 and 2 to their respective data points.

REFERENCES

1. Klotz, Jerome. Markov Chain Clustering of Births by Sex. Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume IV: Biology and Health. University of California Press, 1972.

2. ______. Statistical Inference in Bernoulli Trials with Dependence. The Annals of Statistics, Volume 1, Number 2, pp. 373-379, 1973.

3. Ralston, Anthony. International Series in Pure and Applied Mathematics. A First Course in Numerical Analysis. McGraw-Hill Book Company, 1965, pp. 332, 334, 337, 343-344, 349-350, 373-376.

Distribution List

Organization	Number of Copies
Defense Documentation Center Cameron Station Alexandria, Virginia 22314	12
Director US Army Ballistic Research Laboratory ATTN: DRDAR-TSB-S (STINFO) Aberdeen Proving Ground, Maryland 21005	1
Director US Army TRADOC Systems Analysis Activity ATTN: ATAA-SL (Tech Lib) White Sands Missile Range, New Mexico 88002	1
Mr Gerald Anderson US Army Materiel and Readiness Command 5001 Eisenhower Avenue Alexandria, Virginia 22333	1
Mr James Brown US Army Materiel Systems Analysis Activity Aberdeen Proving Ground, Maryland 21005	1
Mr George Schecter Suite 540 7956 Jones Branch Drive McLean, Virginia 22101	1
Deputy Commander US Army Combined Arms Combat Developments Activity ATTN: ATCA-CAA-O Fort Leavenworth, Kansas 66027	10

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER TP 4-78	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) 6 REGRESSION FOR MARKOV BERNOULLI RANDOM VARIABLES.		5. TYPE OF REPORT & PERIOD COVERED 9 Final rept.,
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) 10 Edmund H./Inselmann, PhD		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Combat Operations Analysis Directorate US Army Combined Arms Combat Development Activity Fort Leavenworth, Kansas 66027		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS ACN 22273
11. CONTROLLING OFFICE NAME AND ADDRESS Combat Operations Analysis Directorate US Army Combined Arms Combat Development Activity Fort Leavenworth, Kansas 66027		12. REPORT DATE 11 April 1978
		13. NUMBER OF PAGES 10
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) 12 11p.		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

14 CACDA-TP-4-78

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Missile simulations, hit probability, Markov chain, Bernoulli random variable

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper deals with the estimation of the regression coefficients when the data are sequences of Bernoulli random variables that form Markov chains. The method used is an extension of Klotz's papers.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED LB

408 081

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)